

KAUSHALENDRA SINGH

Full Stack AI Engineer - LLM Systems • Agentic AI • Backend Architecture
+91 6398519072 | yadavkausha4a5@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EXPERIENCE

FaxLab

Frontend Developer Intern

Remote
Feb 2026 – Mar 2026

- Built UI for AI learning platform serving **500K+** users using Next.js and TypeScript.
- Integrated LLM-powered feedback systems; improved performance by **30%** via SSR, async processing, and optimized pipelines.
- Instrumented structured logging and error tracking for real-time AI-driven content delivery APIs.

GRS Worker

Full Stack Developer Intern

Remote
Sep 2025 – Nov 2025

- Built scalable backend using Node.js, Prisma, PostgreSQL with Redis caching and secure authentication.
- Reduced data errors by **40%** using Zod validation and structured service-layer workflows.
- Integrated Twilio + QR systems; containerized and deployed all services using Docker with CI/CD pipelines.

OPEN SOURCE

Joplin (54K+ GitHub Stars)

[Contributions](#)

- Fixed core issues across desktop, mobile, and sync engine (React, Electron, Node.js); delivered **merged and approved PRs**.
- Collaborated directly with maintainers on a large-scale real-world codebase; followed structured PR review workflows.

PROJECTS

Buildify Labs - Agentic AI Business Intelligence

[Live](#) | [GitHub](#)

Next.js, Node.js, LangChain, LangGraph, OpenAI, Pinecone, Zustand

- Built agentic AI platform using **LangGraph** (ReAct pattern) converting natural language into dynamic dashboards.
- Designed **RAG pipeline**: document embeddings → Pinecone vector search → structured JSON → real-time UI rendering.
- Implemented **short-term + long-term memory** with session state and context window management for multi-step queries.

MyAttendance

[Live](#) | [GitHub](#)

Next.js, MongoDB, Redis, PWA

- Built PWA used by **1300+** users with role-based access control and optimized MongoDB queries for high concurrency.
- Implemented Redis caching layer reducing DB load; added structured error logging for production monitoring.

Postly - Real-time Social App

[Live](#) | [GitHub](#)

Node.js, Socket.io, MongoDB, Kafka

- Built real-time event-driven backend using Socket.io + Kafka message queue; implemented JWT auth and async data processing.
- Designed modular service-layer architecture with scalable API design patterns and live data update subscriptions.

CosmoDetect - Object Detection API — Hackathon Winner

[Live API](#) | [GitHub](#)

Python, YOLOv11, FastAPI, PyTorch, OpenCV

- Built and deployed object detection inference API using PyTorch + FastAPI; optimized preprocessing pipeline with OpenCV + PIL.
- Reduced inference latency via batched preprocessing and async request handling; documented API with auto-generated Swagger docs.

AI EXPERIENCE

- **Agentic Systems:** LangChain, LangGraph; ReAct & Plan-Execute patterns; tool calling; multi-step reasoning workflows.
- **RAG & Retrieval:** Embedding models, Pinecone/FAISS vector DBs, semantic search, context & memory management.
- **LLM Systems:** OpenAI/Gemini APIs with structured outputs, prompt engineering, state handling in AI workflows.
- **Computer Vision:** YOLOv11, PyTorch, OpenCV, PIL — deployed FastAPI inference services with latency optimization.

EDUCATION

Harcourt Butler Technical University, Kanpur

Bachelor of Technology

2023 – 2027

TECHNICAL SKILLS

Languages: JavaScript/TypeScript, Python, SQL, C++

Frontend: Next.js, React, Tailwind, Zustand

Backend: Node.js, Express, FastAPI, Prisma, REST/GraphQL, Kafka, Redis

AI/ML: LangChain, LangGraph, OpenAI, Gemini, PyTorch, YOLOv11, FAISS, Pinecone

Database: MongoDB, PostgreSQL, MySQL

DevOps: Docker, AWS, CI/CD, Logging & Monitoring